Sequence and phylogenetic analysis of the complete mitogenome of *Myzus persicae* (Hemiptera: Aphididae)

YANG Hong[1,2], WANG Jia-Jia[2], DAI Ren-Huai[2,*], YU Hui-Ping[3], YANG Mao-Fa[1,2]

(1. College of Tobacco Science, Guizhou University, Guiyang 551700, China; 2. Entomology Institute, Guizhou University, Guiyang 550025, China; 3. Bijie Branch of Guizhou Tobacco Corporation, Bijie, Guizhou 551700, China)

Abstract: 【Aim】 Aphididae insects are extremely polyphagous agricultural pests. The present study aims to better understand the phylogenetic relationship within Aphididae by mitogenome analysis. 【Methods】 The complete mitogenome of *Myzus persicae* (Hemiptera: Aphididae) was sequenced by next-generation sequencing (NGS) and PCR amplification, and sequence analysis in comparison with other Aphididae insects was made. The Bayesian and Maximum Likelihood phylogenetic analyses based on the 13 PCG sequences were performed. 【Results】 The mitogenome of *M. persicae* (GenBank accession no. KU_236024) is 17 832 bp in size and harbors an A + T content of 84.1%, an AT-skew of 0.094 and a GC-skew of −0.296. It contains 13 protein-coding genes (PCGs) (*cox*1 – 3, *nad*1 – 6, *nad4L*, *atp*6, *atp*8 and *cytb*), 22 tRNA genes, two rRNA genes (*rrnL* and *rrnS*) and two long non-coding regions. The gene order and orientation of the mitogenome of *M. persicae* are similar to those of other aphids. All protein-coding genes (PCGs) start with an ATN initiation codon and terminate with a TAA stop codon, except for *nad*4, which terminates with a single T nucleotide as an incomplete stop codon. In the mitogenome of *M. persicae*, there is a non-coding region (307 bp) between $tRNA^{Glu}$ and $tRNA^{Phe}$, which includes variable numbers of tandem repeats in a lineage-specific manner. The mitogenome of *M. persicae* has the control region of 2 531 bp which is the longest in all the sequenced aphid mitogenomes. In the mitogenome of *M. persicae*, six domains with 44 helices are present in the secondary structure of *rrnL* and three domains with 24 helices are present in the secondary structure of *rrnS*. Phylogenetic analysis based on 13 PCGs from *M. persicae* and 20 other insect species showed that every superfamily in the tree forms a monophyletic clade. 【Conclusion】 Within the Aphidinae subfamily, the monophyly of Aphidini and Macrosiphini had statistically high values. *M. persicae* was clustered with *D. noxia*, and *C. salicicola* was positioned most basally within the clade of Macrosiphini. The results of this study provide some valuable molecular data for reconstructing the phylogenetic relationship in Aphididae.

Key words: Aphid; *Myzus persicae*; mitogenome; RNA; secondary structure; phylogeny

1 INTRODUCTION

In general, insect mitochondrial genomes (mitogenomes) are double-stranded DNA molecules that are 15 – 20 kb in length; and contain 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes, and a control region (Boore, 1999; Cameron, 2014; Wang *et al.*, 2015b). Insect gene orders are mostly identical to the inferred ancestral arrangement, but some species show gene rearrangements (Boore, 1999; Dowton *et al.*, 2002; Thao *et al.*, 2004; Lee *et al.*, 2009; Cameron, 2014). At present, more than 113 complete hemipteran mitogenomes have been sequenced and annotated (Wang *et al.*, 2015b; Li *et al.*, 2016). Mitogenomes are usually used in biogeographic, molecular, systematic, and population studies (Hua *et al.*, 2008; Li *et al.*, 2012; Ma *et al.*, 2012).

Myzus persicae (Sulzer) (Hemiptera: Aphididae) is an extremely polyphagous agricultural pest, feeding on more than 40 families of plants including many major crops such as beans, potatoes, tobaccos, and so on, causing direct damage to plants and indirect damage by the non-persistent transmission of important viruses (Nikolakakis *et al.*, 2003). Aphididae contains more than 5 000 known species, and shows considerable genetic variation with regards to host-plant adaptation (Li *et al.*, 2012; Wang *et al.*, 2013; Wang *et al.*, 2015b). Until now, 11 mitogenomes of Aphididae have been sequenced (Table 1). In the present study, we

基金项目: 贵州省烟草公司毕节市公司科技项目"毕节市烟蚜捕食性天敌生防体系的构建与应用"([2016]82); 贵州省烟草公司科技项目"烟蚜优效捕食性天敌应用技术研究"([2015]11); 贵州省科技厅项目"贵州省烟草品质遗传改良与生物转化科技创新人才团队建设"[黔科合人才团队(2015)4011]

作者简介: 杨洪, 男, 1975 年生, 贵州六枝人, 博士, 教授, 主要从事农业昆虫与害虫防治研究, E-mail: agr.hyang@gzu.edu.cn

* 通讯作者 Corresponding author, E-mail: rhdai69@163.com

收稿日期 Received: 2016-08-10; 接受日期 Accepted: 2016-12-14

Table 1 The mitogenomes of insects used in this study

Species	Family	Size (bp)	GenBank accession no.	References
Myzus persicae *	Aphididea	17 382	KU_236024	This study
Acyrthosiphon pisum *	Aphididea	16 971	NC_011594	IAGC, 2010
Aphis gossypii *	Aphididea	15 869	NC_024581	Zhang *et al.*, 2016
Aphis glycines	Aphididea	12 529	KC_840675	Wang *et al.*, 2013
Cavariella salicicola *	Aphididae	16 317	NC_022682	Wang *et al.*, 2013
Cervaphis quercus *	Aphididea	15 272	NC_024926	Wang *et al.*, 2014
Diuraphis noxia *	Aphididea	15 784	NC_022727	Wang *et al.*, 2014
Hormaphis betulae	Aphididea	15 088	NC_029495	Li *et al.*, 2015
Pterocomma pilosum	Aphididea	13 002	KC_840676	Wang *et al.*, 2013
Schizaphis graminum *	Aphididea	15 721	NC_006158	Thao *et al.*, 2004
Sitobion avenae *	Aphididea	15 180	NC_024683	Zhang *et al.*, 2014
Mindarus keteleerfoliae	Aphididae	18 199	KP_722576	Wang *et al.*, 2015a
Aleurochiton aceris *	Aleyrodidea	15 388	NC_006160	Thao *et al.*, 2004
Aleurodicus dugesii *	Aleyrodidea	15 723	NC_005939	Thao *et al.*, 2004
Bemisia afer *	Aleyrodidea	14 968	NC_024056	Wang *et al.*, 2016
Bemisia tabaci *	Aleyrodidea	15 322	NC_006279	Thao *et al.*, 2004
Neomaskellia andropogonis *	Aleyrodidea	14 496	NC_006159	Thao *et al.*, 2004
Tetraleurodes acaciae *	Aleyrodidea	15 080	NC_006292	Thao *et al.*, 2004
Trialeurodes vaporariorum *	Aleyrodidea	18 414	NC_006280	Thao *et al.*, 2004
Cacopsylla coccinea *	Psyllidea	14 832	NC_027087	Que *et al.*, 2015
Pachypsylla venusta *	Psyllidea	14 711	NC_006157	Thao *et al.*, 2004
Paratrioza sinica *	Psyllidea	14 863	NC_024577	Zhang *et al.*, 2014
Abidama producta	Cercopidae	15 277	NC_015799	Liu *et al.*, 2014
Aquarius paludum	Gerroidea	15 380	NC_012841	Hua *et al.*, 2009
Coptosoma bifaria	Pentatomoidea	16 179	NC_012449	Hua *et al.*, 2008
Nilaparvata muiri	Delphacidae	14 371	NC_024627	Unpublished

* Species used in phylogenetic analysis in this study.

sequenced and annotated the complete mitogenome of *M. persicae*, which typifies the subfamily Macrosiphini.

2 MATERIALS AND METHODS

2.1 Experimental sample and genomic DNA extraction

The specimens of *M. persicae* used in this study were obtained from a laboratory stock reared on tobacco in June 2015. The voucher specimens are deposited in Institute of Entomology, Guizhou University, Guiyang, China (GZU-HO-000011). They were washed twice by vortexing in absolute alcohol and dried out at room temperature before DNA extraction. Total DNA was isolated using a DNeasy Tissue Kit (QIAGEN, Germany), according to the manufacturer's protocols. The DNA was sequenced by next-generation sequencing (NGS) (Illumina HisSeq 4000 and 2Gb raw data, Berry Genomic, Beijing, China) and PCR amplification.

2.2 Assembling sequence, PCR amplification, cloning, and sequencing

The sequence data were assembled by Geneious R9 (Kearse *et al.*, 2012). When necessary, species-specific primers (Table 2) were designed based on known sequence by the program Primer 6.0 (Premier Biosoft, Palo Alto, CA, USA). The fragments were amplified with Taq DNA polymerase (TaKaRa, Dalian, China) under the following conditions: 95℃ for 3 min; 35 cycles of 95℃ for 30 s, 48 – 55℃ (depending on primer combination) for 30 s and 70℃ for 2 min (depending on putative length of the fragments); and 70℃ for 10 min. All the PCR fragments were cloned into the pMD19-T sequencing vector (TaKaRa) and sequenced by Invitrogen Biotech (Beijing, China) and Sangon Biotech (Shanghai, China). Each fragment was assembled into a contig and then assembled into the complete mtDNA by SeqMan (DNASTAR Inc., Madison, WI, USA).

Table 2 PCR primers used in this study

Primers	Primer sequence (5′-3′)	Annealing temperature (℃)	Fragment size (bp)
MP-F1	AAAAGAAGCTGCTAACTATC	48	550
MP-R1	GGTATTTGAAGTTGGTTGAA		
MP-F2	TTCAACCAACTTCAAATACC	48	500
MP-R2	ATCATTATGTGGTTTTCCTT		
MP-F3	GAAGAAACAGGAGTAGGTGC	50	900
MP-R3	TATATCTTTTCCTCCTTCTT		
MP-F4	GATAACCCCAACCATAAATT	52	800
MP-R4	AAGTTTGATCTAATGGGTGA		
MP-F5	ATAATAGGGTATCTAATCCTAGTTT	48	3 000
MP-R5	CAACTACAAACGTCATGCCT		

2.3 Sequence analysis and gene annotation

Thirteen PCGs, two ribosomal RNA genes and the control region of the mitogenome of *M. persicae* were identified by comparison with the mitogenome sequences of other Aphididae insects. The nucleotide sequences of 13 PCGs were translated into amino acid sequences on the basis of the invertebrate mitochondrial genetic code. The relative synonymous codon usage (RSCU) values and A + T content of nucleotide sequence were calculated using MEGA 6.0 (Tamura *et al.*, 2013). The composition skewness was calculated according to the formula: AT skew = ; A − T]/[A + T], GC skew = [G − C]/[G + C] (Perna and Kocher, 1995). The tRNA genes and secondary structure were identified using tRNA-SE v. 1.21 (Schattner *et al.*, 2005) and ARWEN (Laslett and Canbäck, 2008), and drawn using DNASIS v. 2.5 (Hitachi Engineering, Tokyo, Japan). The secondary structures of the two rRNA genes (*rrnL* and *rrnS*) were predicted after the *Cavariella salicicola* (Wang *et al.*, 2013) and drawn by RNA structure (Reuter and Mathews, 2010).

2.4 Phylogenetic analysis

Eighteen mitogenomes of Aphididae, Aleyrodidae and Psyllidae species (marked with an asterisk in Table 1) downloaded from the GenBank and the mitogenome of *M. persicae* were used to analyze the phylogenetic relationship among Sternorrhyncha, and *Nilaparvata muiri*, *Aquarius paludum*, *Coptosoma bifaria* and *Abidama producta* were used as outgroups. The sequences of 13 PCGs were aligned by MEGA v. 6.06 (Tamura *et al.*, 2013) and the exported data were formatted by Clustal X (Thompson *et al.*, 1997) for phylogenetic analysis. Phylogenetic analysis was performed using Bayesian inference (BI) analysis with MrBayes v. 3.1.2 (Huelsenbeck and Ronquist, 2001) under the TVM + I + G model and Maximum Likelihood with MEGA v6.0. The model was chosen with Modetest 3.7 (Nylander *et al.*, 2004). BI analysis, two simultaneous runs of 1 000 000 generations were conducted for the matrix. The "temperature" parameter was set to a default value of 0.2. Each one was sampled every 1 000 generations with a burn-in of 25%. ML tree was evaluated through MEGA 6.06 (Tamura *et al.*, 2013) using the test of bootstrap method with 1 000 replications under Kimura 2-parameter model.

3 RESULTS

3.1 Genome structures and base composition

The complete mtDNA of *M. persicae* (GenBank accession no. KU_236024) is 17 832 bp in size (Fig. 1). The mitogenome contains 13 PCGs (*cox*1-3, *nad*1-6, *nad4L*, *atp*6, *atp*8 and *cytb*), 22 tRNA genes, two rRNA genes (*rrnL* and *rrnS*). All genes (Table 3) are similar to those described in other hemipteran insects (Wang *et al.*, 2015b).

The nucleotide composition of the whole mitogenome of *M. persicae* is as follows: A, 46.0%; T, 38.1%; G, 5.6%; and C, 10.3%. This is similar to that of other insects. The nucleotide composition, AT-skew, and GC-skew are shown in Table 4.

3.2 Protein-coding genes

The mitogenome of *M. persicae* includes 13 PCGs, Which is similar to other Aphididae insects (Wang *et al.*, 2013, 2015a; Zhang *et al.*, 2014; Zhang *et al.*, 2016). Nine of the 13 PCGs (*nad*2, *cox*1, *cox*2, *atp*8, *atp*6, *cox*3, *nad*3, *nad*6, and *cytb*) are coded on the majority strand (J-strand), four PCGs are coded on the minority strand (N-strand) (Table 3 and Fig. 1). The start codon of the PCGs is a typical ATN, five (*nad*2, *cox*1, *cox*2, *atp*8, and *nad*4) starting with ATA, five (*atp*6, *nad*3, *nad*5, *nad*6, and *nad*1) with ATT, and the remainder with ATG. Alternative start codons have

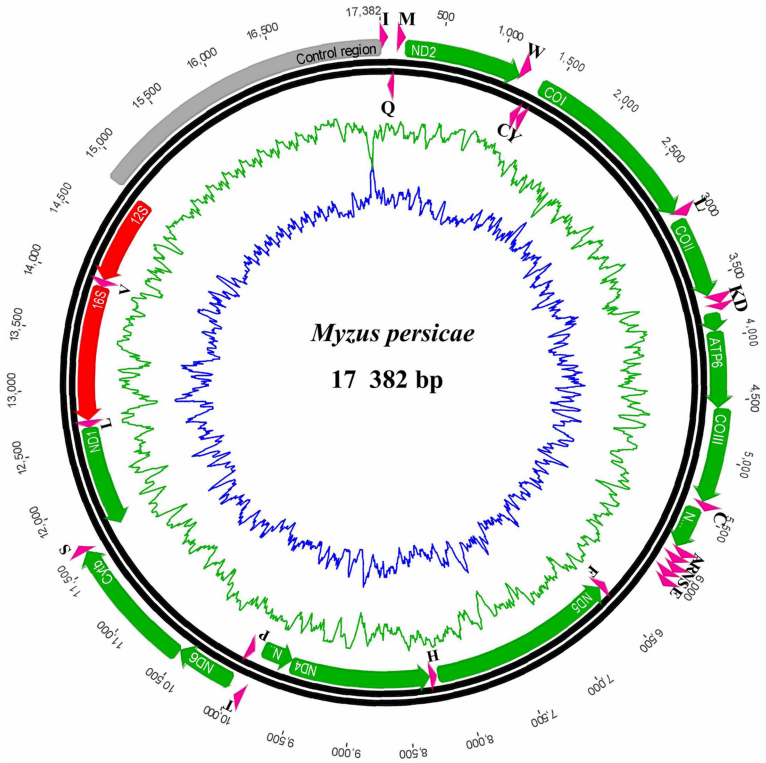

Fig. 1　Map of the mitogenome of *Myzus persicae*
Graphic representation of AT (green) and GC (blue) contents and their variation throughout the genome.

been found in other insects, such as TTG in *Hackeriella veitchi* (Lutz-Bonengel *et al.*, 2004), GTG in *Vargula hilgendorfii* (Ogoh and Ohmiya, 2004), ATAA in *Drosophila melanogaster* (Ballard, 2000), and so on. All PCGs terminate with the stop codon TAA, except for *nad*4 which uses a single T nucleotide as an incomplete stop codon. A partial stop codon can usually occur in insects (Wang *et al.*, 2015b).

3.3　Codon usage

The RSCU values of the mitogenome of *M. persicae* are summarized in Table 5. The 13 PCGs in the mitogenome of *M. persicae* consist of 3 648 codons in total. The codon GCG is absent in these PCGs. Five frequent codons are UUU (F), UUA (L), AUU (I), AUA (I), and UAU (Y). These codons are all composed of A and U nucleotides, indicating a high usage of A and T in *M. persicae* PCGs. This is similar in other Aphididae insects (Wang *et al.*, 2015a).

3.4　Transfer and ribosomal RNA genes

The mitogenome of *M. persicae* contains a typical set of 22 tRNAs, and all of them were predicted by tRNA-SE (Schattner *et al.*, 2005) and ARWEN (Laslett and Canbäck, 2008). The tRNA genes range from 62 bp ($tRNA^{Val}$) to 73 bp ($tRNA^{Lys}$) in size. Fourteen tRNA genes are coded on the J-strand and the others on the N-strand. The secondary structures of the *M. persicae* tRNAs are shown in Fig. 2. All tRNA genes are folded into a typical clover-leaf secondary structure, except $tRNA^{Ser}$ (AGN) which lacks a DHU arm and this is a common feature in hemipteran insects.

The ranges of rRNA genes were identified by alignment with other Aphididae species. The large ribosomal gene (*rrnL*) and small ribosomal gene (*rrnS*) of *M. persicae* are located between $tRNA^{Leu}$ (CUA) and $tRNA^{Val}$ and between $tRNA^{Val}$ and the A + T-rich region (Fig. 1). The *rrnL* and *rrnS* genes are 1 260 and 812 bp in length, respectively, With an AT content of 84.9% (Table 4). The secondary structures of *rrnL* and *rrnS* were predicted based on the secondary structure models from other insects (Gillespie *et al.*, 2006; Wang *et al.*, 2013; Ma *et al.*, 2016). In the mitogenome of *M. persicae*, six domains with 44 helices are present in the secondary structure of *rrnL* (Fig. 3), which is very similar with that in *C. salicicola*, the first reported in Aphididae (Wang *et al.*, 2013). However, some differences exist between different aphid groups, for example, the secondary structure of *rrnL* of *M. persicae* has a bigger loop (22 bp) than *C. salicicola* (5 bp) in helix H2077. Three domains with 24 helices are present in the secondary structure of *rrnS* in the mitogenome of *M. persicae* (Fig. 4), and H673 is slightly different in *C. salicicola*.

Table 3 The organization of the mitogenome of *Myzus persicae*

Gene	Direction	Position	Size (bp)	Anti/start codon	Stop codon	Intergenic nucleotides
tRNA$^{Ile\ (I)}$	J	1 – 65	65	GAT	–	
tRNA$^{Gln\ (Q)}$	N	63 – 128	66	TTG	–	11
tRNA$^{Met\ (M)}$	J	140 – 206	67	CAT	–	0
nad2	J	207 – 1 184	978	ATA	TAA	– 3
tRNA$^{Trp\ (W)}$	J	1 182 – 1 249	68	TCA	–	– 9
tRNAC$^{ys\ (C)}$	N	1 241 – 1 308	68	GCA	–	4
tRNA$^{Tyr\ (Y)}$	N	1 313 – 1 378	66	TGT	–	1
cox1	J	1 380 – 2 915	1 536	ATA	TAA	– 5
tRNA$^{Leu-1\ (L)}$	J	2 911 – 2 978	68	TAA	–	3
cox2	J	2 982 – 3 653	672	ATA	TAA	2
tRNA$^{Lys\ (K)}$	J	3 656 – 3 728	73	CTT	–	0
tRNA$^{Asp\ (D)}$	J	3 729 – 3 791	63	GTC	–	9
atp8	J	3 801 – 3 950	150	ATA	TAA	1
atp6	J	3 952 – 4 584	633	ATT	TAA	– 1
cox3	J	4 584 – 5 369	786	ATG	TAA	– 1
tRNA$^{Gly\ (G)}$	J	5 369 – 5 433	65	TCC	–	0
nad3	J	5 434 – 5 787	354	ATT	TAA	– 1
tRNA$^{Ala\ (A)}$	J	5 787 – 5 850	64	TGC	–	– 1
tRNA$^{Arg\ (R)}$	J	5 850 – 5 915	66	TCG	–	0
tRNA$^{Asn\ (N)}$	J	5 916 – 5 981	66	GTT	–	– 2
tRNA$^{Ser\ (S)}$	J	5 980 – 6 045	66	GCT	–	2
tRNA$^{Glu\ (E)}$	J	6 048 – 6 111	64	TTC	–	– 6
Repeat region	–	6 106 – 6 412	307	–	–	
tRNA$^{Phe\ (F)}$	N	6 413 – 6 477	65	GAA	–	0
nad5	N	6 478 – 8 217	1 740	ATT	TAA	5
tRNA$^{His\ (H)}$	N	8 223 – 8 285	63	GTG	–	0
nad4	N	8 286 – 9 603	1 318	ATA	T	– 1
nad4L	N	9 603 – 9 896	294	ATG	TAA	– 2
tRNA$^{Thr\ (T)}$	J	9 895 – 9 956	62	TGT	–	1
tRNA$^{Pro\ (P)}$	N	9 958 – 10 023	66	TGG	–	1
nad6	J	10 025 – 10 519	495	ATT	TAA	– 1
cytb	J	10 519 – 11 637	1 119	ATG	TAA	4
tRNA$^{Ser-2\ (S)}$	J	11 642 – 11 706	65	TGA	–	10
nad1	N	11 717 – 12 652	936	ATT	TAA	0
tRNA$^{Leu-2\ (L)}$	N	12 653 – 12 717	65	TAG	–	0
rrnL	N	12 718 – 13 977	1 260	–	–	0
tRNA$^{Val\ (V)}$	N	13 978 – 14 039	62	TAC	–	0
rrnS	N	14 040 – 14 851	812	–	–	0
Control region	–	14 852 – 17 382	2 531	–	–	0

J: Majority strand (J-strand); N: Minority strand (N-strand).

Table 4 Nucleotide composition and skewness of the mitogenome of *Myzus persicae*

Feature	T (%)	C (%)	A (%)	G (%)	A + T (%)	AT-skew	GC-skew
Whole genome	38.1	10.3	46.0	5.6	84.1	0.094	– 0.296
Protein-coding genes	47.6	8.9	35.2	8.3	82.8	– 0.150	– 0.035
1st codon	37.0	9.9	47.3	6.1	84.3	0.122	– 0.228
2nd codon	39.0	10.8	44.3	5.7	83.3	0.064	– 0.305
3rd codon	39.0	10.0	46.5	4.9	85.5	0.088	– 0.352
tRNA genes	40.1	8.4	45.5	6.0	85.6	0.063	– 0.167
rRNA genes	38.8	10.1	46.1	5.0	84.9	0.086	– 0.338
Control region	38.5	9.9	48.7	2.9	87.2	0.117	– 0.547

Table 5 The number of codons and relative synonymous codon usage (RSCU) in *Myzus persicae*

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	428	1.71	UCU(S)	34	1.43	UAU(Y)	212	1.72
UUC(F)	72	0.29	UCC(S)	2	0.08	UAC(Y)	34	0.28
UUA(L)	380	3.83	UCA(S)	87	3.65	UAA(*)	12	0.36
UUG(L)	49	0.49	UCG(S)	2	0.08	UAG(*)	–	–
CUU(L)	76	0.77	CCU(P)	28	1.53	CAU(H)	53	1.61
CUC(L)	12	0.12	CCC(P)	7	0.38	CAC(H)	13	0.39
CUA(L)	60	0.61	CCA(P)	37	2.03	CAA(Q)	55	1.62
CUG(L)	18	0.18	CCG(P)	1	0.05	CAG(Q)	13	0.38
AUU(I)	370	1.69	ACU(T)	30	1.28	AAU(N)	176	1.77
AUC(I)	40	0.18	ACC(T)	4	0.17	AAC(N)	23	0.23
AUA(I)	247	1.13	ACA(T)	59	2.51	AAA(K)	123	1.80
AUG(M)	30	1.00	ACG(T)	1	0.04	AAG(K)	14	0.20
GUU(V)	78	2.33	GCU(A)	18	1.38	GAU(D)	61	1.74
GUC(V)	5	0.15	GCC(A)	6	0.46	GAC(D)	9	0.26
GUA(V)	39	1.16	GCA(A)	28	2.15	GAA(E)	70	1.52
GUG(V)	12	0.36	GCG(A)	0	0	GAG(E)	22	0.48
UGU(C)	25	1.85	CGC(R)	2	0.20	AGA(R)	33	3.30
UGC(C)	2	0.15	CGA(R)	14	1.40	AGG(R)	6	0.60
UGA(*)	–	–	CGG(R)	1	0.10	GGU(G)	23	1.35
GGG(G)	8	0.47						

3.5 Non-coding regions

The mitogenome of *M. persicae* contains 2 882 bp of non-coding DNA (Table 3). The intergenic spacer between $tRNA^{Ser}$(UCN) and *nad*1 is 10 bp in length, which is the same as those of *Acyrthosiphon pisum*, *Diuraphis noxia*, and *Schizaphis graminum*. This region corresponds to one of the binding sites of a mitochondrial transcription termination factor (Roberti *et al.*, 2003), and possesses a 7 bp conserved motif (Cameron and Whiting, 2008; Zhang *et al.*, 2014). The corresponding motif is ATACTAA in *M. persicae*, *A. pisum*, and *D. noxia*. Generally, insect mitogenomes contain a control region and some short intergenic spacers, however, there is a special repeat region between $tRNA^{Phe}$ and $tRNA^{Glu}$ in the mitogenome of *M. persicae*, as found in other Aphidinae insects except *Mindarus keteleerifoliae*.

The repeat region is located between $tRNA^{Glu}$ and $tRNA^{Phe}$. This region consists of multiple tandem repeat units, followed by a partial repeat unit. Both the length and copy number of the repeat units are different among aphid mitogenomes (*S. graminum*, 635 bp; *D. noxia*, 644 bp; and *A. pisum*, 1 513 bp). The repeat region (307 bp) in *M. persicae* consists of two repeat units (79 bp for each), with the first overlapping $tRNA^{Glu}$ by 6 bp.

The control region of the mitogenome of *M. persicae* is rich in A + T (87.2%) and located between *rrnS* and $tRNA^{Ile}$, the same as in other aphid species, it can be divided into four parts: repeat regions, A + T-rich zone, a conserved poly (T) stretch, and a stem-loop region (Wang *et al.*, 2013, 2015a; Wang *et al.*, 2014). The lengths of control regions in aphid mitogenomes are variable, for instance, the mitogenome of *M. persicae* (2 531 bp) is currently the largest aphid mitogenomes (from 657 bp in *Cervaphis quercus* to 1 336 bp in *A. pisum*). This difference in length is mainly attributed to their repeated regions. The first repeat is copied nine times, and every unit is 87 bp in length; the second repeat is copied eight times, and every unit is 131 bp in length. Furthermore, a microsatellite "$(TA)_{25}$" was found ahead of the poly (T) motif.

3.6 Phylogenetic analysis

In this study, the amino acid sequences of the 13 PCGs were concatenated to reconstruct phylogenetic relationships among the mitogenome of *M. persicae* and those of 20 other insect species. These mitogenomes were found to represent three families within Hemiptera. The phylogenetic trees generated from ML and BI analyses showed the same topology (Fig. 5). The analyses supported that every

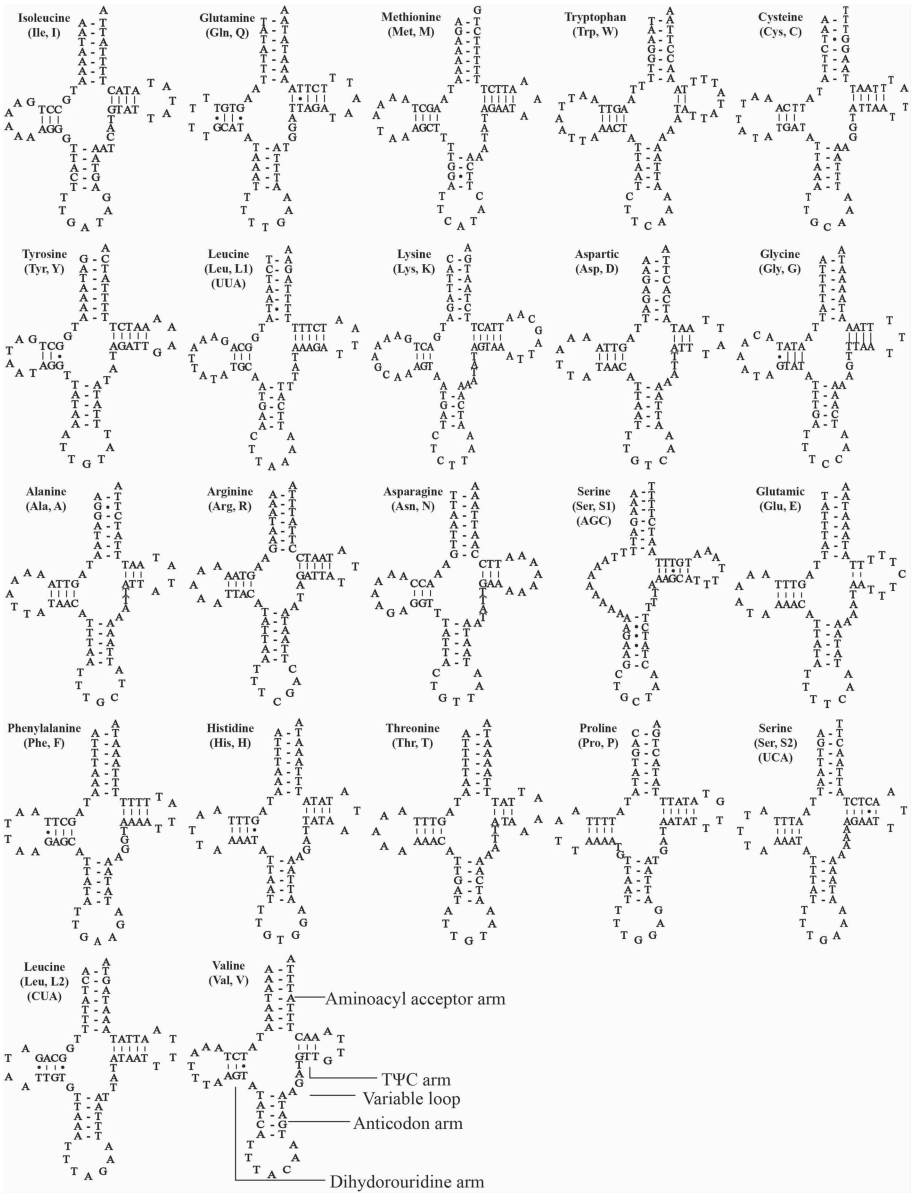

Fig. 2 Inferred secondary structures of 22 tRNA genes in the mitogenome of *Myzus persicae*

The tRNAs are labeled with the abbreviations of their corresponding amino acids. The minus sign indicates Watson-Crick base pairing and a dot indicates G-U base pairing.

superfamily in the tree forms a monophyletic clade. BI analyses showed that every superfamily in the tree forms a monophyletic clade. Aphididae and Aleyrodidae group into one clade and then demonstrate a stable sister relationship with Psyllidae, and according to the taxonomy they all belong to Sternorrhyncha. The monophyly of Macrosiphini and Aphidini was recovered in the analyses and well supported. The mitogenomes of *A. pisum* and *S. avenae* are grouped with *D. noxia*, *M. persicae* and *C. saliciola* in the Macrosiphini and cluster with others. Our phylogenetic analyses of these families were clearly segregated and demonstrated a similar topology with traditional and other molecular data (Blackman and Eastop, 1984; Ferrari *et al.*, 2006; von Dohlen *et al.*, 2006; Wang *et al.*, 2015a).

4 DISCUSSION AND CONCLUSION

The present paper reports the complete mitogenome sequence of *M. persicae* (17 832 bp). The mitogenome of *M. persicae* has similar gene order and orientation to those of other aphids, but has the longest control region of all the sequenced aphid mitogenomes and its RNA structure is similar with that of other insects. In the mitogenome of *M. persicae*, a special repeat region (307 bp) exists between $tRNA^{Phe}$ and $tRNA^{Glu}$, as found in other Aphidinae insects except *M. keteleerifoliae*. Phylogenetic analysis based on 13 PCGs showed that every superfamily in the tree forms a

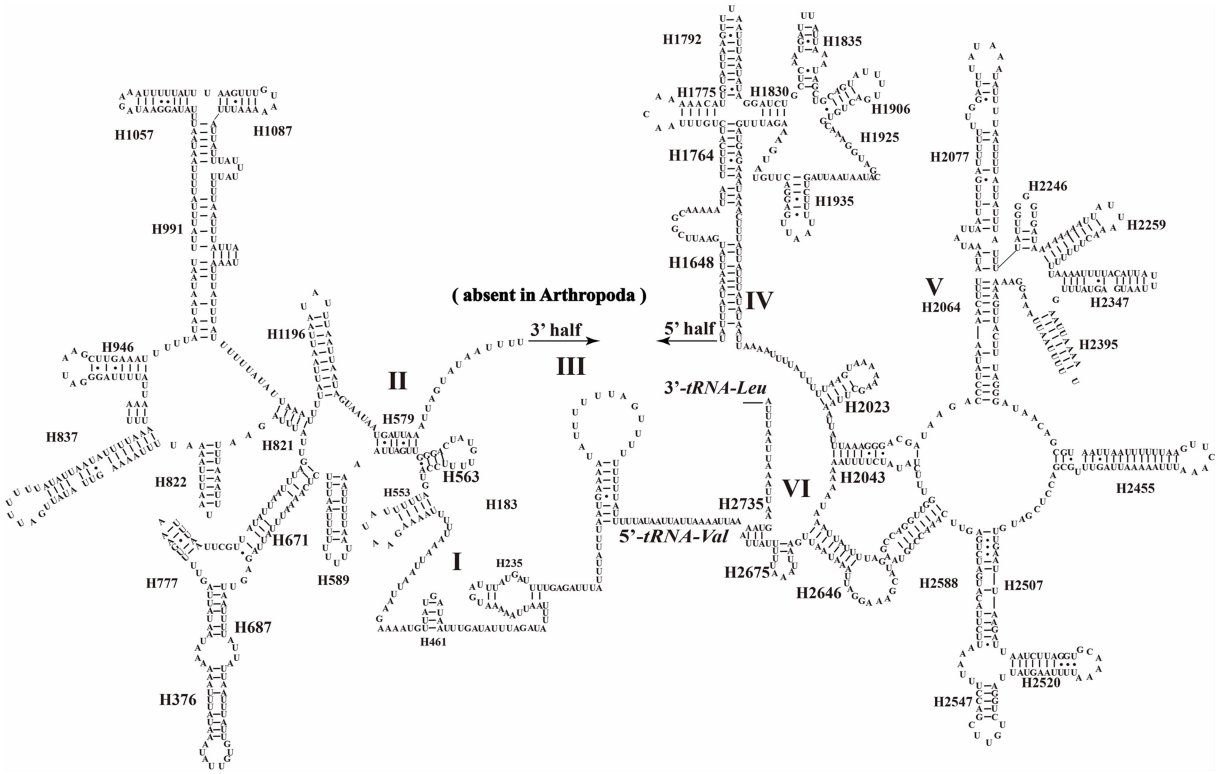

Fig. 3 Predicted secondary structure of the *rrnL* gene in the mitogenome of *Myzus persicae*

Roman numerals represent the conserved domain structures. Dashes indicate Watson-Crick base pairings and dots indicate G-U base pairing.

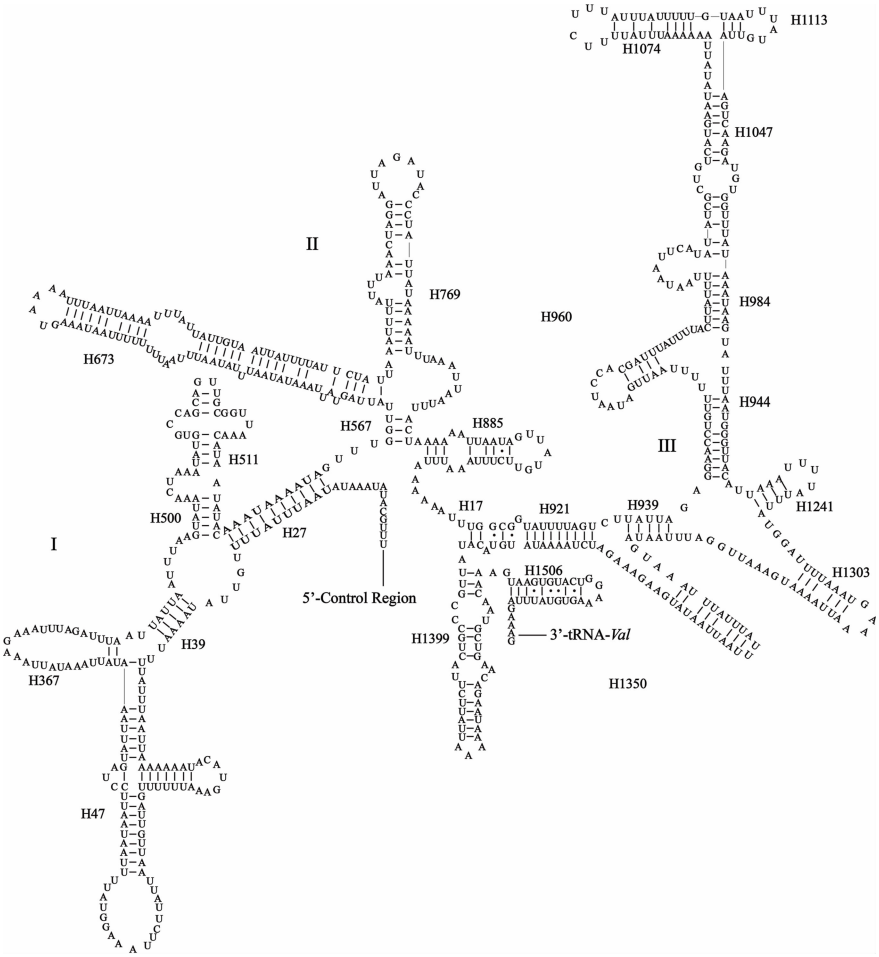

Fig. 4 Predicted secondary structure of the *rrnS* gene in the mitogenome of *Myzus persicae*

Roman numerals represent the conserved domain structures. Dashes indicate Watson-Crick base pairings and dots indicate G-U base pairing.

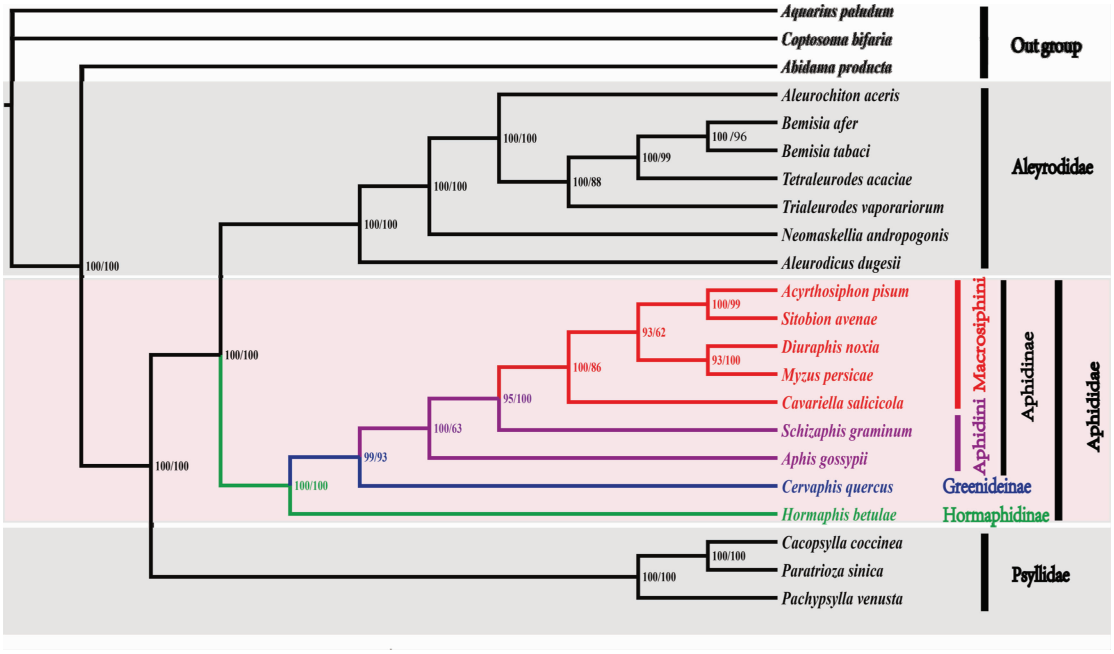

Fig. 5 Phylogenetic tree inferred by Bayesian Inference (BI) and maximum likelihood (ML) methods based on the amino acid sequences of the 13 protein-coding genes (PCGs) from *Myzus persicae* and 21 other insect species

Bayesian posterior probabilities (front) and ML bootstrap support values (back) are shown near the nodes.

monophyletic clade. The mitogenome information inferred in this study may be limited, therefore, future research will involve in more taxon sampling to resolve this problem.

References

Ballard JWO, 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Molec. Evol.*, 51(1): 48 – 63.

Blackman RL, Eastop VF, 1984. Aphids on the World's Crops. An Identification and Information Guide. Wiley, Chichester, UK.

Boore JL, 1999. Animal mitochondrial genomes. *Nucleic Acids Res.*, 27(8): 1767 – 1780.

Cameron SL, 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. *Ann. Rev. Entomol.*, 59: 95 – 117.

Cameron SL, Whiting MF, 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene*, 408(1): 112 – 123.

Dowton M, Castro L, Austin A, 2002. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology'. *Invertebr. Syst.*, 16(3): 345 – 356.

Ferrari J, Godfray HCJ, Faulconbridge AS, Prior K, Via S, 2006. Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution*, 60(8): 1574 – 1584.

Gillespie J, Johnston J, Cannone J, Gutell R, 2006. Characteristics of the nuclear (18S, 5. 8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements. *Insect Mol. Biol.*, 15(5): 657 – 686.

Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W, 2008. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). *BMC Genomics*, 9: 610.

Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W, 2009. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. *BMC Evol. Biol.*, 9: 134.

Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8): 754 – 755.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, 2012. Geneious R9-Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12): 1647 – 1649.

Laslett D, Canbäck B, 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, 24(2): 172 – 175.

Lee W, Kang J, Jung C, Hoelmer K, Lee SH, Lee S, 2009. Complete mitochondrial genome of brown marmorated stink bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and phylogenetic relationships of hemipteran suborders. *Mol. Cells*, 28(3): 155 – 165.

Li H, Liu H, Song F, Shi A, Zhou X, Cai W, 2012. Comparative mitogenomic analysis of damsel bugs representing three tribes in the family Nabidae (Insecta: Hemiptera). *PLoS ONE*, 7(9): e45925.

Li H, Shi A, Song F, Cai W, 2016. Complete mitochondrial genome of the flat bug *Brachyrhynchus hsiaoi* (Hemiptera: Aradidae). *Mitochondr. DNA*, 27(1): 14 – 15.

Li YQ, Chen J, Qiao GX, 2015. Complete mitochondrial genome of the aphid *Hormaphis betulae* (Mordvilko) (Hemiptera: Aphididae: Hormaphidinae). *Mitochondr. DNA*, DOI: 10. 3109/19401736. 2015. 1118071.

Liu J, Bu C, Wipfler B, Liang A, 2014. Comparative analysis of the mitochondrial genomes of Callitettixini spittlebugs (Hemiptera: Cercopidae) confirms the overall high evolutionary speed of the AT-rich region but reveals the presence of short conservative elements at the tribal level. *PLoS ONE*, 9(10): e109140.

Lutz-Bonengel S, Sänger T, Pollak S, Szibor R, 2004. Different methods to determine length heteroplasmy within the mitochondrial control region. *Int. J. Legal. Med.*, 118 (5): 274 – 281.

Ma C, Yang P, Jiang F, Chapuis MP, Shali Y, Sword GA, Kang L, 2012. Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. *Mol. Ecol.*, 21 (17): 4344 – 4358.

Ma L, Dong WW, Jiang GF, Wang X, 2016. The complete mitochondrial genome of *Brachmia macroscopa* (Lepidoptera: Gelechiidae) and its related phylogenetic analysis. *J. Insect Sci.*, 16 (1): 9.

Nikolakakis N, Margaritopoulos J, Tsitsipis J, 2003. Performance of *Myzus persicae* (Hemiptera: Aphididae) clones on different host-plants and their host preference. *Bull. Entomol. Res.*, 93 (03): 235 – 242.

Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey J, 2004. Model test-Bayesian phylogenetic analysis of combined data. *Syst. Biol.*, 53(1): 47 – 67.

Ogoh K, Ohmiya Y, 2004. Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorfii* (Crustacea, Ostracoda) with duplicate control regions. *Gene*, 327(1): 131 – 139.

Perna NT, Kocher TD, 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.*, 41(3): 353 – 358.

Que S, Yu L, Xin T, Zou Z, Hu L, Xia B, 2015. Complete mitochondrial genome of *Cacopsylla coccinae* (Hemiptera: Psyllidae). *Mitochondr. DNA*, 27(5): 3169 – 3170.

Reuter JS, Mathews DH, 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11 (1): 1.

Roberti M, Polosa PL, Bruni F, Musicco C, Gadaleta MN, Cantatore P, 2003. DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA. *Nucleic Acids Res.*, 31(6): 1597 – 1604.

Schattner P, Brooks AN, Lowe TM, 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucl. Acids Res.*, 33(Suppl. 2): W686 – W689.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: molecular evolutionary genetics analysis version 6. 0. *Mol. Biol. Evol.*, 30(12): 2725 – 2729.

Thao ML, Baumann L, Baumann P, 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol. Biol.*, 4(1): 1.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25(24): 4876 – 4882.

von Dohlen CD, Rowe CA, Heie OE, 2006. A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. *Mol. Phylogenet. Evol.*, 38(2): 316 – 329.

Wang HL, Xiao N, Yang J, Wang XW, Colvin J, Liu SS, 2016. The complete mitochondrial genome of *Bemisia afer* (Hemiptera: Aleyrodidae). *Mitochondr. DNA*, 27(1): 98 – 99.

Wang Y, Chen J, Jiang LY, Qiao GX, 2015a. The complete mitochondrial genome of *Mindarus keteleerifoliae* (Insecta: Hemiptera: Aphididae) and comparison with other Aphididae insects. *Int. J. Mol. Sci.*, 16(12): 30091 – 30102.

Wang Y, Chen J, Jiang LY, Qiao GX, 2015b. Hemipteran mitochondrial genomes: features, structures and implications for phylogeny. *Int. J. Mol. Sci.*, 16(6): 12382 – 12404.

Wang Y, Huang XL, Qiao GX, 2013. Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. *PLoS ONE*, 8 (10): e77511.

Wang Y, Huang XL, Qiao GX, 2014. The complete mitochondrial genome of *Cervaphis quercus* (Insecta: Hemiptera: Aphididae: Greenideinae). *Insect Sci.*, 21(3): 278 – 290.

Zhang B, Ma C, Edwards O, Fuller S, Kang L, 2014. The mitochondrial genome of the Russian wheat aphid *Diuraphis noxia*: large repetitive sequences between trnE and trnF in aphids. *Gene*, 533(1): 253 – 260.

Zhang S, Luo J, Wang C, Lv L, Li C, Jiang W, Cui J, Rajput LB, 2016. Complete mitochondrial genome of *Aphis gossypii* Glover (Hemiptera: Aphididae). *Mitochondr. DNA*, 27(2): 854 – 855.

烟蚜线粒体基因组全序列及系统发育分析

杨 洪[1,2]，汪佳佳[2]，戴仁怀[2,*]，喻会平[3]，杨茂发[1,2]

（1. 贵州大学烟草学院，贵阳 550025；2. 贵州大学昆虫研究所，贵阳 550025；
3. 贵州省烟草公司毕节市公司，贵州毕节 551700）

摘要：【目的】蚜虫是杂食性农业害虫。本研究旨在通过线粒体基因组分析更好地了解蚜科昆虫的系统发育关系。【方法】结合第二代测序和 PCR 扩增技术获得了烟蚜 *Myzus persicae* 线粒体基因组全序列，与蚜科其他昆虫进行了对比分析；以贝叶斯法和最大似然法基于 13 个蛋白编码基因对蚜科进行了系统发育分析。【结果】烟蚜线粒体基因组（GenBank 登录号：KU_236024）序列全长 17 832 bp，A + T 含量 84.1%，AT 偏斜为 0.094，GC 偏斜为 −0.296。包含 13 个蛋白编码基因（*cox*1 −3，*nad*1 −6，*nad4L*，*atp*6，*atp*8 和 *cytb*），22 个 tRNA，2 个 rRNA 基因（*rrnL* 和 *rrnS*）和 2 个长的非编码区，其基因排列顺序与已知的蚜科昆虫相似，除了 *nad*4 以单独的 T 结尾，所有的蛋白编码基因均以 ATN 作为起始密码子，TAA 作为终止密码子。在烟蚜线粒体基因组中，$tRNA^{Glu}$ 和 $tRNA^{Phe}$ 中间有一段 307 bp 的非编码区，该编码区包含 2 个重复单元，烟蚜的控制区长 2 531 bp，是所有测序蚜虫线粒体基因组中最长的。*rrnL* 的二级结构包含 6 个结构域，44 个茎环结构；*rrnS* 的二级结构有 3 个结构域，24 个茎环结构。基于烟粉虱和其他 20 种昆虫的 13 个蛋白编码基因重建的 BI 和 ML 系统发育树，与传统形态学分类结果一致。【结论】蚜亚科和长管蚜亚科的单系性得到了很好的支持；在长管蚜亚科的分支中，*M. persicae* 与 *D. noxia* 聚成一支，并且 *C. salicicola* 位于进化枝的底部。本研究结果为蚜科系统发生关系重建积累了有价值的数据资料。

关键词：蚜虫；烟蚜；线粒体基因组；RNA；二级结构；系统发育

中图分类号：Q969　　**文献标识码**：A　　**文章编号**：0454-6296（2017）01-0084-11

（责任编辑：赵利辉）